

OIPE

RAW SEQUENCE LISTING DATE: 02/28/2002 PATENT APPLICATION: US/10/027,801 TIME: 10:23:54

Input Set : N:\Crf3\RULE60\10027801.raw.txt
Output Set: N:\CRF3\02282002\J027801.raw

```
1 <110> APPLICANT: Swanson, Ronald V.
              Feldman, Robert A.
              Schleper, Christa
      4 <120> TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
      5 <130> FILE REFERENCE: DCORP.002A
      6 <140> CURRENT APPLICATION NUMBER: 10/027,801
      7 <141> CURRENT FILING DATE: 2001-12-21
      8 <150> PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020
W--> 9 <151> PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29
     10 <160> NUMBER OF SEQ ID NOS: 123
     11 <170> SOFTWARE: FastSEQ for Windows Version 3.0
     13 <210> SEQ ID NO: 1
     14 <211> LENGTH: 32998
     15 <212> TYPE: DNA
     16 <213> ORGANISM: Cenarchaeum symbiosum
     17 <220> FEATURE:
     18 <221> NAME/KEY: CDS
                                                              ENTER
    19 <222> LOCATION: (7604)...(8908)
     20 <221> NAME/KEY: CDS
    21 <222> LOCATION: (8961)...(9767)
     22 <221> NAME/KEY: CDS
     23 <222> LOCATION: (10545)...(10922)
     24 <221> NAME/KEY: CDS
    25 <222> LOCATION: (13944)...(14612)
    26 <221> NAME/KEY: CDS
    27 <222> LOCATION: (18638)...(20149)
    28 <221> NAME/KEY: CDS
    29 <222> LOCATION: (20554)...(20955)
    30 <221> NAME/KEY: CDS
    31 <222> LOCATION: (20956)...(21834)
    32 <221> NAME/KEY: CDS
    33 <222> LOCATION: (25151)...(26377)
    34 <221> NAME/KEY: CDS
    35 <222> LOCATION: (27535)...(28002)
    36 <221> NAME/KEY: CDS
    37 <222> LOCATION: (28065)...(29483)
    38 <400> SEQUENCE: 1
                                                                                      60
    39
             gatecttgae etetgegett attgeageea tggaetgaee ggeegtgegg ggetaaataa
    40
             agctgaggcg ccgcctgcag gctctgctca gccgttgatt atacagtact cgcactcgca
                                                                                     120
    41
             ggtcttgctt tcgtcatctt ttgccctgca cttggcgtgc tggcccgact ggcactggac
                                                                                     180
    42
                                                                                     240
             gcagatctcg tctatgttgt ctatcgaggt gtactttttg gactggtcaa agccgaagcc
                                                                                     300
    43
             gccgtccttt ggtccaacca tgaagcggtc agggccgtta tgccttaatt atctttgccg
```

teggggeggg geogettetg eeggegggag eeceeettga ggeegeteee eggetegttg

360

RAW SEQUENCE LISTING DATE: 02/28/2002 PATENT APPLICATION: US/10/027,801 TIME: 10:23:54

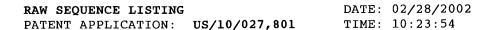
45			_		ctatgattgc		420
46					tgcccatgcg		480
47					ggacagacag		540
48					cacccccac		600
49					ctgggcgggg		660
50					tgagaacaga		720
51	caactgcccg	atcgtttttc	aggtacctac	tgggcttttt	ggagttcgtt	gatcaaaagc	780
52	ggtacaggta	ttctatacat	gccagtcttg	gctggaaaaa	taaattgaag	atcggcggat	840
53	cctatctacg	agcgcctgcc	tgcttttcac	ttgactagcc	ggagtacttc	gtctgcaatt	900
54	tctgtatagc	taggcattta	tgtagttgag	atacatgtcc	gcgggatccg	tttcatagtc	960
55	tgaatcaaac	accggatcat	tccttctctt	taattcctta	aatgcctgat	gcagttcaag	1020
56	caggacggtt	acatcatcgc	gtttgattat	ccgttctgtt	gtttcagctt	tttgctcatt	1080
57	ctcatccatg	attataggct	acggtatttg	actaaaaagg	tttccatctt	catgagtcgt	1140
58	gtgtttgccg	tggaattacc	taccggggga	acataaaaaa	atgagtcata	aacagcgcac	1200
59	tgcatccacc	gtggacccat	gagacaacga	gccggcagcg	gtgcacagca	ccgagtacaa	1260
60					tccagacgga		1320
61	-	_		•	ccgtggtgtt		1380
62					gactggagta		1440
63					aaattcttgt		1500
64					cgatatgcgc		1560
65	-	_			ttttgacttt		1620
66			_	-	tacaagttcc		1680
67					ttgccctaac		1740
68					tgcttcttga		1800
69					tggaataata		1860
70	-		•		attcgatgag		1920
71					gacattttta		1980
72					cccggactg		2040
73					aaatccatac		2100
74		_	_		tatccagttg		2160
75		-			atagtcgtga		2220
7 <i>5</i> 76		_					2280
					gtcgtttgct		2340
77					tcataccata		
78					catgcagttc		2400
79					gcttttggtc		2460
80			_		tgattttaat		2520
81					acaggcagag		2580
82		-			tccgctgggg		2640
83					accggtgttt		2700
84	, ,,	•		, ,,,	accaccttcc		2760
85	-				gtctcgggat		2820
86					ccggcctgcc		2880
87					ctaagagcga		2940
88					cgacggtcta		3000
89	•				ctgctgcagg		3060
90					gagctctcgc		3120
91					ccaatagtgg		3180
92	gatcgctaag	ccagactttc	gtctatgaat	tccgtgcgtt	tggaaatcca	ttcagtctag	3240
93	tttttggctt	tgccctcttc	agcggatttc	tgacccgctt	gaactaaact	ttgggcccct	3300

RAW SEQUENCE LISTING DATE: 02/28/2002 PATENT APPLICATION: US/10/027,801 TIME: 10:23:54

0.4	++~~+~+	++	+~~~~~~	~~~~~	aanaat aana	at = t = = = = = =	3360
94 95			tgccgcccca		-		3420
96			ctgcaggaaa				3420
97	_		tgactcccag ccacggggtc		_		3540
98		-					3600
99	-		caccgggttg				3660
100	_		tacccgacaa		-		3720
100						accggcaccg	3780
						tttttattaa	
102						cagctgcagg	3840
103			caggactaat				3900
104						gaacttgcag	3960
105						ctaacgcaaa	4020
106	-		-	-		cgaacggttc	4080
107			cccctatcc			_	4140
108		_				ggcagtcctt	4200
109						ttacggtggt	4260
110		-				aatcggtcca	4320
111			tcgtcccaat				4380
112			tcggtactct				4440
113						agcttacctc	4500
114						ttggggacct	4560
115			cccctctcgg				4620
116			tccgttcgga				4680
117			ctaccggaaa				4740
118			agcgccagtc		-	_	4800
119			cagaactgct			_	4860
120	_		cgactggctt	_		-	4920
121	_		tgctgcgaga			-	4980
122			gcaagctccc				5040
123			actttcagct			-	5100
124			tctgtaagca				5160
125			cctcacggta				5220
126			ccaatcttcg				5280
127			cttcgtctag		_		5340
128	_		ctggggccat				5400
129			gtttgggctc				5460
130	_		tcgtggtact		_		5520
131			gattcctatt				5580
132			gcttgccacg				5640
133			cggcatattc				5700
134			gggggccgct				5760
135	ttgcaccatg	caaagatcat	gtgcattctg	ttcaaaccag	tttctaagga	ggtgatccga	5820
136	ccgcaggttc	ccctacggtc	accttgttac	gacttttccc	ttgtcgctta	cctcaagttc	5880
137			acctcactaa				5940
138			attcactgcg				6000
139			cagccctcag				6060
140	tcctcctttc	ggatatggaa	cccattgtca	ctaccattgc	agcccgcgtg	tggccccaga	6120
141			gccgtggccc				6180
142	ccgctaattc	gccccactgc	tccggagagc	aatggtggca	actagaggca	aggatctcgc	6240

## RAW SEQUENCE LISTING DATE: 02/28/2002 PATENT APPLICATION: US/10/027,801 TIME: 10:23:54

143	tcqttacctq	acttaacagg	acatctcacg	gcacgagctg	gcgacggcca	tgcaccacct	6300
144	ctcagcttgt	ctggtagagt	cttcagcttg	accttcacac	tgctgtctct	ccgggtaaga	6360
145	tttctqqcqt	tgactccaat	tgaaccgcag	gcttcacccc	ttgtggtgct	ccccgccaa	6420
146	ttcctttaag	tatcatactt	gcgtacgtac	ttcccaggcg	gcaaacttaa	cggcttccct	6480
147	geggeactge	actggctctt	acgccaatgc	atcactgagt	ttgcattgtt	tacagctggg	6540
148	actacccggg	tatctaatcc	ggtttgctcc	cccagctttc	atccctcacc	gtcggacgtg	6600
149	ttctagtaga	ccgccttcgc	cacagggggt	catcgataga	tcagaggatt	ttaccccttc	6660
150	ctaccgagta	ccgtctacct	ctcccactcc	ctagccgtgc	agtatttccg	gcagcctatg	6720
151	cattgagcac	atagatttaa	ccgaaaactt	acacqqcaqq	ctacggatgc	tttaggccca	6780
152	ataateetee	tgaccacttg	aggtgctggt	tttaccgcgg	cggctgacac	cagaacttgc	6840
153	ccacccctta	ttcgccggtg	gttttaagac	cqqtaaaaqa	tttctttagc	agaaaacact	6900
154	cagattaacc	ttgtcgtgct	ttcgcacatt	gcaaagtttt	ctcqcctqct	gcgccccata	6960
155	agacctagat	ccgtgtctca	gtacccatct	ccgagcctct	cctctcagag	cccgtatctg	7020
156	ttatageett	ggtgggccat	tacctcacca	acaagctgat	agaccgcagt	cccatcctac	7080
157	ggggataaat	catttgggcc	acaaaccatt	ccaggcatag	tagcctatca	gatattattc	7140
158	tcactttccc	gaggttatcc	ccatccatag	gttagattga	ctacqtqtta	ctgagccgtc	7200
159	tagattatat	tgctacaatg	actcgcatgg	cttagtatca	atccgatage	agtcaggtcc	7260
	aggaggatga	accggattca	taattggatt	atttttttt	tgttaagtac	gcttgtactt	7320
160	ttaassttas	acagaatgca	cataatette	acatctcaga	tatgaccctt	cgatcatacc	7380
161	ctggaactga	gtgcgtaact	adaddccadc	gaatcacaat	atggtacaat	accatgcatt	7440
162	cicattetgt	gccgctcttg	ggaggccagc	gateggateg	acceptecat	gggcatataa	7500
163	categeaage	atttccgccc	acadagaaca	castcadad	ccagatctac	ctgtatgatg	7560
164	accategeeg	ctgattaaat	tatagagaga	acaacctact	accatagate	tagaacacaa	7620
165	gegateegee	cigatiaaat	catagagaga	geggeetgee	aggtcgaaaa	agtaccacgt	7680
166	gtacagggca	aagaccggcg	tangetteta	gateteteget	coatttataa	caaggtccgc	7740
167	cggcggggtc	agccacaaca	naggitteta	gaageegtat	gtagactact	ggatggggga	7800
168	gagcggcaag	cacctcgtcg	acgreggacgg	gaacaagcac	togggagtag	aggacygggaa	7860
169	ctggagcctg	atactggggc	aegegeegge	gecagecagg	atgastatat	aggggcaget	7920
170	tegeegegge	tggatccacg	ggaccgtcaa	cyaycayacy	acgaaccccc	cadadaccat	7980
171	acgcggcgcg	gtaagcgtgg	cagaaaagac	aaggtacgtc	acguegggga	cggaggccgc	8040
172	catgtatgcg	gcaaggctgg	egegegegea	Lacygycaya	adaacaacag	catatastat	8100
173	cggcggctgg	cacgggtacg	cgtcggggct	gercaagreg	tacattagge	agaagatgt	8160
174	gcccgagagc	ggggggctcg	tegacgaaga	geactetata	ttagastaga	tantantage	8220
175	tgaaggttcc	ctggatgttc	ttgggcgcgc	aggegaegae	Liggeargeg	raggastaga	8280
176	gccgctgctg	ggcggcggcg	gctgcatacc	ggeggatgag	gaetatetge	geggeataca	8340
177	ggagtttgtg	cattcaaggg	gegegetget	tgtcctcgac	gagatagtga	tagggtteeg	8400
178	gtttaggttt	ggctgcgcgt	atgctgcagc	agggctggac	cccgatatag	Lggegetegg	8460
179	caagatagtc	gggggcggat	tccccatagg	ggtgatatgc	ggcaaggacg	aggigalgga	8520
180	aatctccaac	actatatcgc	atgcaaagtc	cgacagggcg	tacatcggcg	geggeacatt	
181	ctctgcaaac	cccgccacga	tgacagcggg	cgcggcagcg	ctcggggagc	tcaaaaagag	8580
182	aaagggcaca	atatacccga	ggataaactc	catgggggac	gacgcaaggg	acaagetete	8640
183	aaagatattt	gggaacaggg	tatccgtgac	cggaaggggc	tcgctgttca	tgactcactt	8700
184	tgttcaagat	ggcgccggca	gggtctcaaa	tgctgcagat	gcggcagcct	gcgatgttga	8760
185	gctgctgcac	aggtaccacc	tggacatgat	cacccgggac	ggcatattct	ttctgccggg	8820
186	caagctgggg	gccatatcgg	cggcgcactc	aaaggccgac	ctcaagacca	tgtattccgc	8880
187	atcagagcgc	tttgcagaag	gcctatgagg	tatagcgccg	gaggaaactt	tgattatacg	8940
188	ggcgtgctgc	cccggggccc	atgatactct	tcggcaagag	cgaccccgcc	gagctggtgc	9000
189	gccaggcgga	cctcctgtgc	agcaagaacc	agttcagggc	ggcaataggc	ctgtacggga	9060
190	aaatcctcaa	ggacgacccg	cagaacaggg	gcgtcctgca	caaaaagggg	ctggcccaga	9120
191	acagggcaaa	aaagtactct	gatgcgatca	cgtgctttga	ccggctgctc	gagcttgaca	9180



192	acaaggacgc	gcccgcgtac	aacaacaagg	ccatagccca	ggccgagctc	ggagacacgg	9240
193	catccgcgct	ggaaaactac	ggcagggcca	tcgaggccga	cccgcggtac	gcgccggcgc	9300
194		ggccgtgctg					9360
195	tcgacagggc	agccgagctg	gaccgacgca	agccgaaccc	gaggttctac	aaggggatag	9420
196	tgctcggcaa	gatgggcagg	cacgaagagg	cgctggcctg	cttcaagggc	gtgtgcaaga	9480
197		ccacgccgac					9540
198	ggcacgccga	ggccctcggg	gagcttgcat	cactgcccgc	ggagcaccgc	gagaacgcca	9600
199	atgtattgta	tgccagggcg	cgcagcctct	cgggccttgg	cagggaggac	gaatccatag	9660
200		aaaggcggcc					9720
201	aaaaggcctt	tgacggaata	cgggacgatc	ccggttcaaa	aagatagccg	gctagaggat	9780
202	cttttttctt	gccgcgtcaa	tccgcatcat	gcggaccttt	tttttgggcc	ccacaagtcg	9840
203	cgattcatag	actggtacat	agaccacctc	caccgccttt	gcggcaaact	cctcccgcag	9900
204	gtcgcgcatg	ccgtcaggcg	ggggcccgcg	cagcttctct	tttagttttg	agagcgcctc	9960
205	ttctgtctcc	acctcggggc	tccgcacatt	ctctgacgca	tcgagtatcc	tccgcgggta	10020
206	cggctccacc	gcgccgggcc	ccgtcttgta	gggaaagtcc	gtctcgccgc	cgtgccggtc	10080
207	aaggcacatc	atcccttctg	attccgcaaa	gacatgctct	tctagctcga	ggtcgaccct	10140
208		agccctgccg					10200
209	gacgccgtcg	cctagcacca	cctcgatcac	gttctggtcc	accttgatcg	ggtgaaccgc	10260
210	ctttctgaaa	aaatccgcag	agtacctggc	ggagacccgg	atcagcgcct	cgtggaccag	10320
211	ctttacagaa	tgcacatgga	cgtcttcttt	ccgcggggcc	ctcataaggg	ccctaaaggc	10380
212	acccgtcttc	tttgcctcta	tcatggcccg	agccgactcc	tcagtcatgg	cgttccgcag	10440
213		ctggtctttc					10500
214	aacgcaaggc	aaggtaataa	tagcctgccg	tctgtaacgg	ccgtatgagg	tcggagggca	10560
215	ggcccggata	catcgaaaag	ttcctaaaga	gggcggacaa	ggcgatagac	aatgcagtcg	10620
216	agcagggcgt	caagagggca	gacgagatac	tagatgacgc	agtcgagctc	ggcaagatca	10680
217	ccgtgggcga	ggcgcaaaaa	agaagcgatg	tgctgctcaa	gcaggccgag	cgggagagca	10740
218	agcggctcaa	gtcaaggggc	gccaaaaagc	tcgaaaaggg	cataggggcg	gcaaaaaaga	10800
219	tggcagccgg	caagggcgac	gcgctagaga	ccctggcaaa	gctcggcgag	ctgagaaagg	10860
220	cggggatcat	aacggagaag	gagtttcgcg	ccaagaaaaa	gaagcttctc	gcggagatct	10920
221	gacttgaagc	cgctagacta	tacccgggac	ggctcgataa	aggaggtcac	aaagaggtgg	10980
222	tttataggca	cgccgtccct	tgtcgacctt	gcaggcgagc	tcggcatatc	tgagagcaac	11040
223	atattccacg	tgacatttcc	cgacggcgca	aagaccaccc	tgcatacgca	cgagggcggg	11100
224	cagctgctca	tagtgacctc	gggcaccggc	agcatgtcaa	tatttgaaaa	gaccggcgga	11160
225	ggcgaggcgg	aatttgcaat	aaaagagaca	gacaggatac	cgctaaagca	gggcagcatc	11220
226	cagtacatac	ctgccggcgt	gctcacgtgc	acggcgcaac	agacggcacc	accctgtccc	11280
227	atatagcggt	aaactacccg	tcgccatcgg	gaaaggagcc	gtatacatta	tggtatgaat	11340
228	ccgactttgc	cagccgggtc	accggcgtgc	tgtaaattat	attatttgag	cctctccagt	11400
229	atcgacaggc	ttacaaggtt	ggtcatcgtt	atccccttgc	ggatcacttc	ccttctagtc	11460
230	ttttcgcagt	acttgttgac	gctctggtgg	ctcttttcgc	tggatacctc	gagcaccaca	11520
231		tgtacgggaa					11580
232	cctattcccg	ccctgcctat	cttctccttt	cttgccacta	ccgacgccac	ctcgcgtatg	11640
233	tectectegg	aatctccgta	ttccagatag	tacatggata	catagctcat	cccgggggat	11700
234	tccctttcga	atatctcctc	gtccatgctg	aataaatagg	aagggccgcc	cccgcggccc	11760
235	tccaccgcct	ttatcatttt	ggggccgttt	ttgaaccttg	ccagcaggta	gtggctggcc	11820
236	-	agtacggctt	-				11880
237		tcgacgagga					11940
238	tatggcgtgc	aggcatatcc	ccccggggag	gcctccgtct	tggacatgca	tgggatccgc	12000
239		aatatctagt					12060
240	gccccggggc	aggcgtgcgt	ggatccatgc	gatagttatt	taaaactagg	atgccgatca	12120

VERIFICATION SUMMARY

DATE: 02/28/2002

PATENT APPLICATION: US/10/027,801

TIME: 10:23:55

Input Set : N:\Crf3\RULE60\10027801.raw.txt Output Set: N:\CRF3\02282002\J027801.raw

L:9 M:256 W: Invalid Numeric Header Field, Wrong Prior FILING DATE:YYYY-MM-DD

 $L:7288 \ M:341 \ W: \ (46) \ "n" \ or "Xaa" \ used, for SEQ ID#:120$